Identify Residential Neighborhoods for families with children in the city of Chicago

Lekshmi R Nair

Introduction/Business Problem

Chicago is one among the biggest cities in the United States. The city is also reported as one among the best big cities to live, with a ranking of 26th out of 62, according to the report by WalletHub, a personal finance website. According to the 2018 global livability index ,Chicago ranks sixth among the top ten best cities to live in the United States. At the same time, there are also concerns regarding the high crime rate in the city, much higher than the US national average. This makes the city an apt choice to explore the residential neighborhoods , that are suitable for families with children .

Given the above background, the purpose of this project is to identify the residential neighborhoods in Chicago city. The target audience is families with children that are planning to settle in the city of Chicago.

Data

We will start by scraping the Wikipedia page https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago. A sample of the table is given below.

Table1:List of Neighborhoods in Chicago

Neighborhood	Community area
Albany Park	Albany Park
Altgeld Gardens	Riverdale
Andersonville	Edgewater
Archer Heights	Archer Heights
Armour Square	Armour Square
Ashburn	Ashburn
Ashburn Estates	Ashburn
Auburn Gresham	Auburn Gresham
Avalon Park	Avalon Park
Avondale	Avondale
Avondale Gardens	Irving Park
Back of the Yards	New City

Belmont Central	Belmont Cragin
Belmont Gardens	Hermosa
Belmont Heights	Dunning

In this page, the neighborhoods are listed according to the community area. This is converted into a data frame. Then the latitude and longitude of each neighborhood are found out using Google map. The sample of the table with the latitude and longitude of each neighborhood is given below.

Table2: Geographical Coordinates

Community area	Latitude	Longitude
Albany Park	41.95431	-87.7213
Archer Heights	41.83793	-87.7133
Armour Square	41.81254	-87.6324
Ashburn	41.76885	-87.7263
Auburn Gresham	41.71191	-87.6209
Austin	41.85774	-87.7141
Avalon Park	41.76634	-87.6153
Avondale	41.95431	-87.7213
Belmont Cragin	41.95341	-87.7641
Beverly	41.68926	-87.6741
Bridgeport	41.81254	-87.6324
Brighton Park	41.82996	-87.6715
Burnside	41.70356	-87.5983
Calumet Heights	41.7414	-87.5546
Chatham	41.76634	-87.6153
Chicago Lawn	41.7953	-87.6961
Clearing	41.77642	-87.8059
Douglas	41.81107	-87.6168
Dunning	41.97283	-87.888
East Garfield Park	41.90081	-87.7147
East Side	41.66634	-87.5594
Edgewater	41.98675	-87.6632
Edison Park	42.00806	-87.8118

Source:Google Map

Foursquare API is used to explore the neighborhood venue categories using the latitude and longitude. Sample data is given below.

Table 3:Venue Categories using API

Community area	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
0	ALBANY PARK	41.95431	-87.72127	Independence Park	41.953624	-87.72369	Park
1	ALBANY PARK	41.95431	-87.72127	The Original Alps	41.953695	-87.719085	Breakfast Spot
2	ALBANY PARK	41.95431	-87.72127	Wing Chong Restaurant	41.95547	-87.721463	Asian Restaurant
3	ALBANY PARK	41.95431	-87.72127	Hot Nails	41.953941	-87.717524	Nail Salon
4	ALBANY PARK	41.95431	-87.72127	Tony's Finer Foods	41.956154	-87.722428	Supermarket
5	ALBANY PARK	41.95431	-87.72127	Cafe Urbano	41.956872	-87.723879	Café
6	ALBANY PARK	41.95431	-87.72127	Hearth & Crust	41.953986	-87.718319	New American Restaurant
7	ALBANY PARK	41.95431	-87.72127	Chicken Works & Salad Company	41.953893	-87.719881	American Restaurant
8	ALBANY PARK	41.95431	-87.72127	El Llano Restaurant	41.953985	-87.719637	Latin American Restaurant

We will get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. The socio economic data of Chicago collected from the Chicago Data Portal is also used here to segment the venue categories for identifying the best residential neighborhoods apt for families and children. Among these indicators, the sample data for the count of elementary, middle and high schools per community area is shown in table4.

Table4: Number of Elementary, Middle and High schools per Community area

	COMMUNITY_AREA_NAME	Number of public schools
0	ALBANY PARK	8
1	ARCHER HEIGHTS	2
2	ARMOUR SQUARE	3
3	ASHBURN	8
4	AUBURN GRESHAM	10

Crowded Housing Rate per Community area

We will enrich the dataset by adding crowded housing rate per community area. Crowded housing rate is defined as the percentage of occupied housing units with more than one person per room among all occupied housing units. The data is taken from the Chicago data portal, based on the US Census data 2016.Sample data is given below.

Table 5:Crowded Housing Rate 2016

Community Area	Percent of Crowded house
Rogers Park	7.7
West Ridge	7.8
Uptown	3.8
Lincoln Square	3.4
North Center	0.3

CRIME DATA PER COMMUNITY AREA

Now the violent crimes by neighborhoods in Chicago for 2018 is imported from renthop.com. Sample data is given below.

Community Area	Violent Crimes	Violent crime percapita
Forest Glen	50	0.0027
Mount Greenwood	72	0.0038
North Center	157	0.0045
Dunning	261	0.0061
Jeffeson Park	187	0.0069

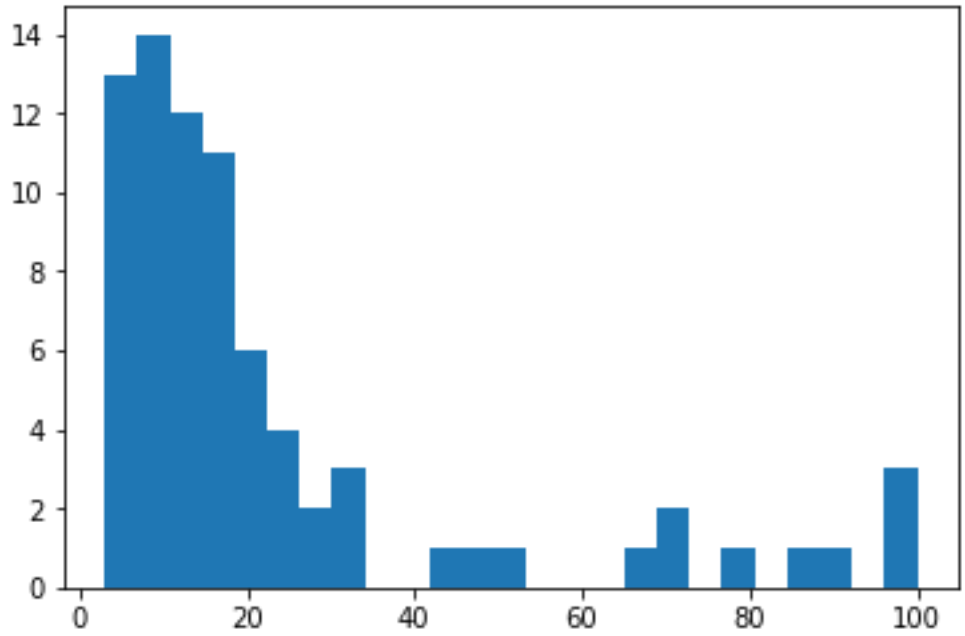
3. Methodology

First, all the data was collected and saved in data frames. The Wikipedia page was scraped for a list of community areas and their corresponding neighborhoods. Google maps are used to find the latitude and longitude of each community area.

Next, Four Square API is used to explore the neighborhoods. The geographical coordinates of each neighborhood were passed to the API which returned a maximum of 100 venue categories and within the fixed 500 meters radius. The resulting data frame consisted of all neighborhood lists by community area with added venues and venue categories.

The number of venue categories in each neighborhood was found out. A histogram was drawn to find out the neighborhoods with the least number of venue categories. The result is shown below.

Figure1: Histogram of the Venue Categories



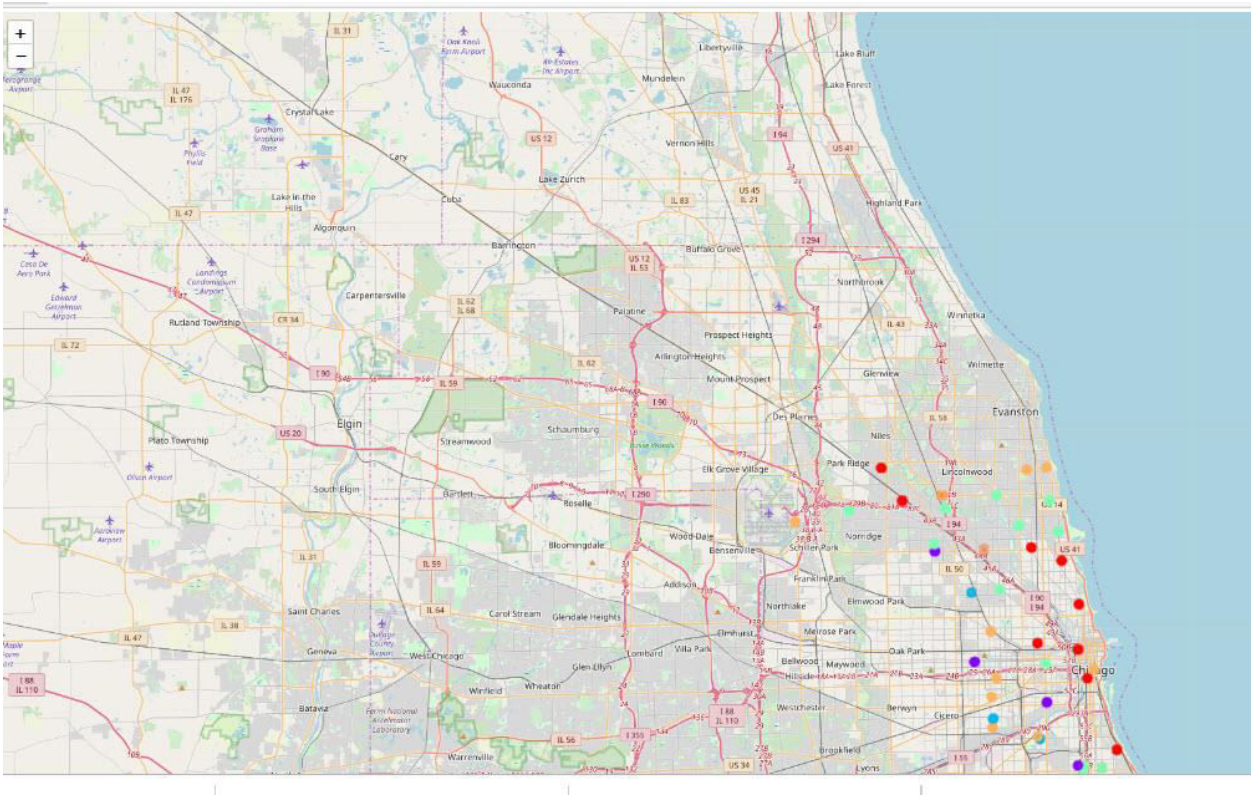
The above histogram shows that many community areas have only a few venue categories (the leftmost bar on the diagram is much higher than the others). These community areas have too little data to make a meaningful analysis; therefore we will exclude them from the dataset.

After excluding the data, the dataset was prepared for analysis using one hot encoding .This was done to pivot venue categories from rows to columns .The mean of frequency of occurrence of each category was calculated.

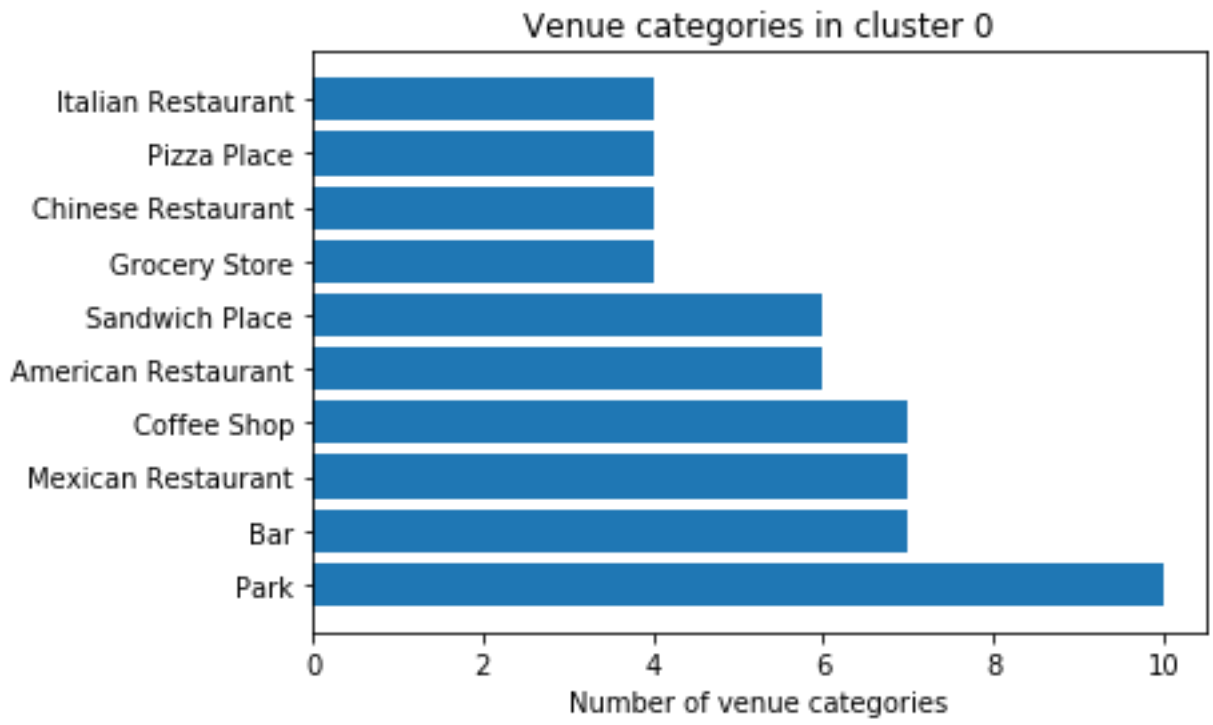
K-means clustering algorithm was used on the above data frame to derive clusters of neighborhoods by community area using 5 as the number of clusters.

Results

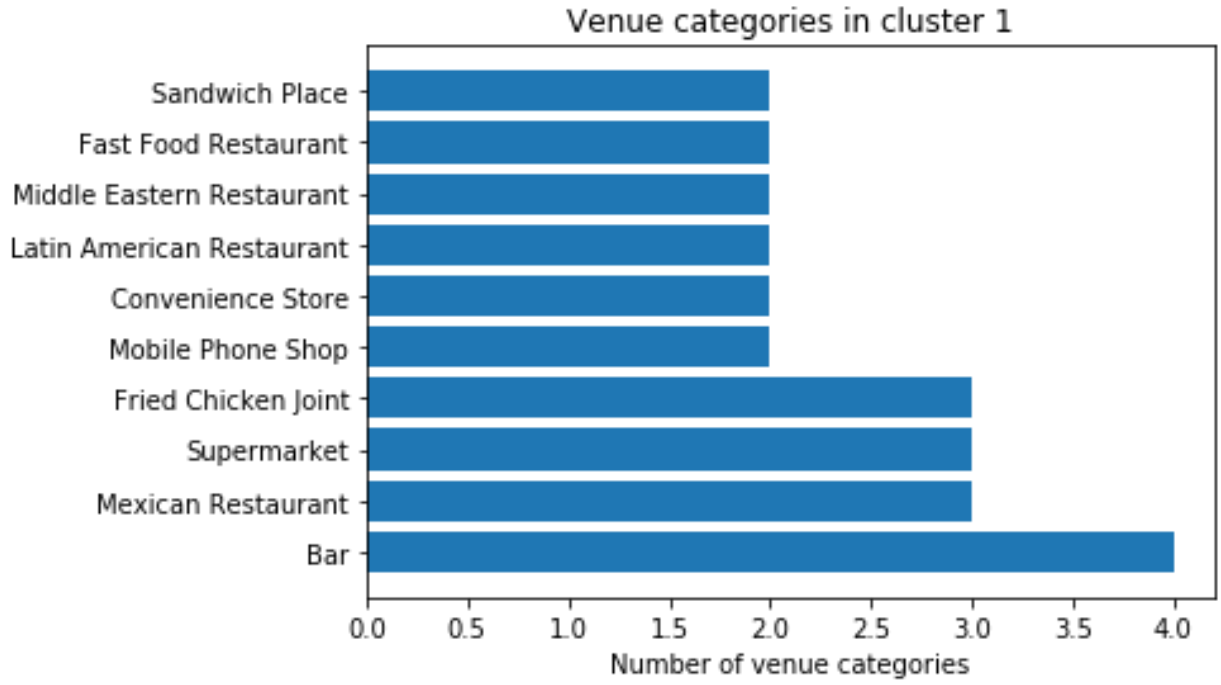
The clusters were visualized on a map as given below.



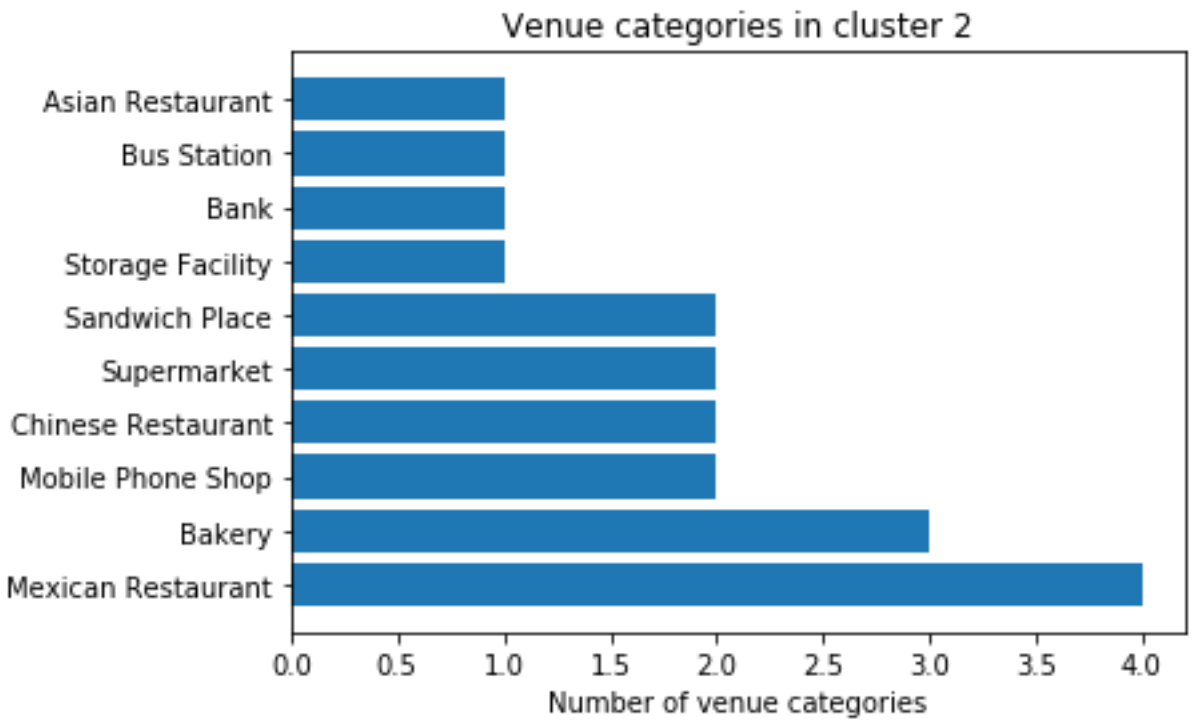
Each cluster was then explored to determine the common venue categories that define each cluster. Results are shown below.



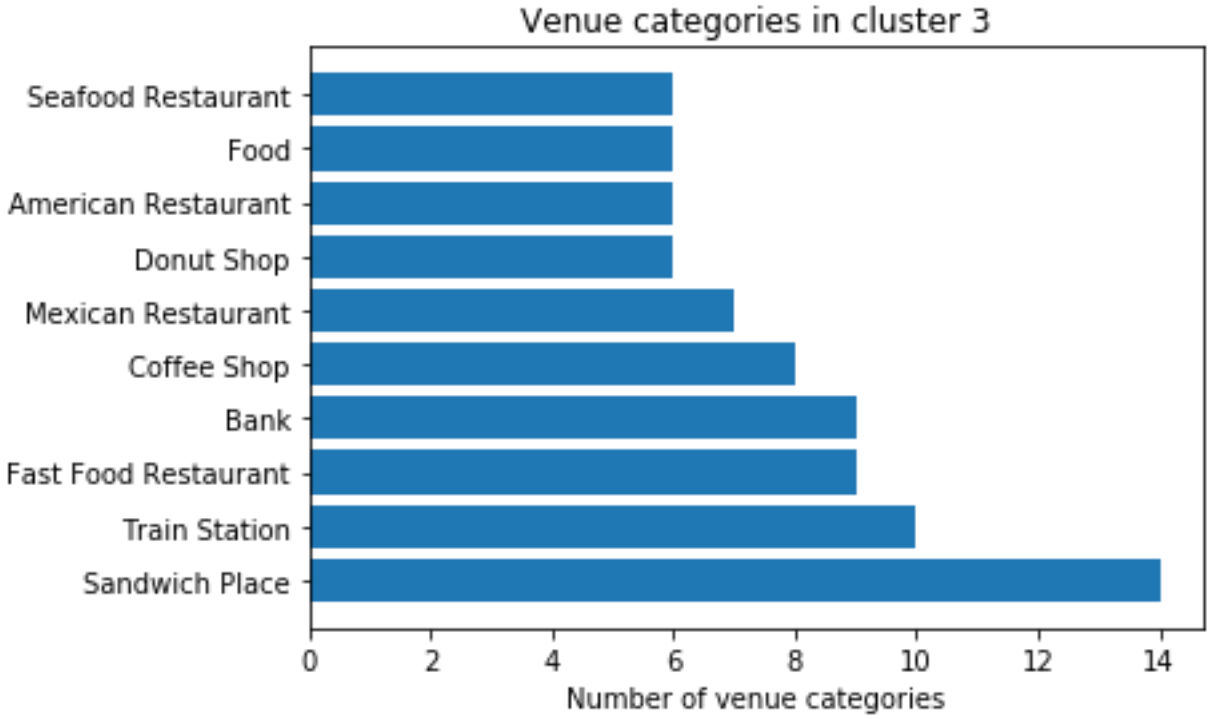
Cluster 0 is one of neighborhoods with parks, bars, coffee shops, restaurants and grocery store.



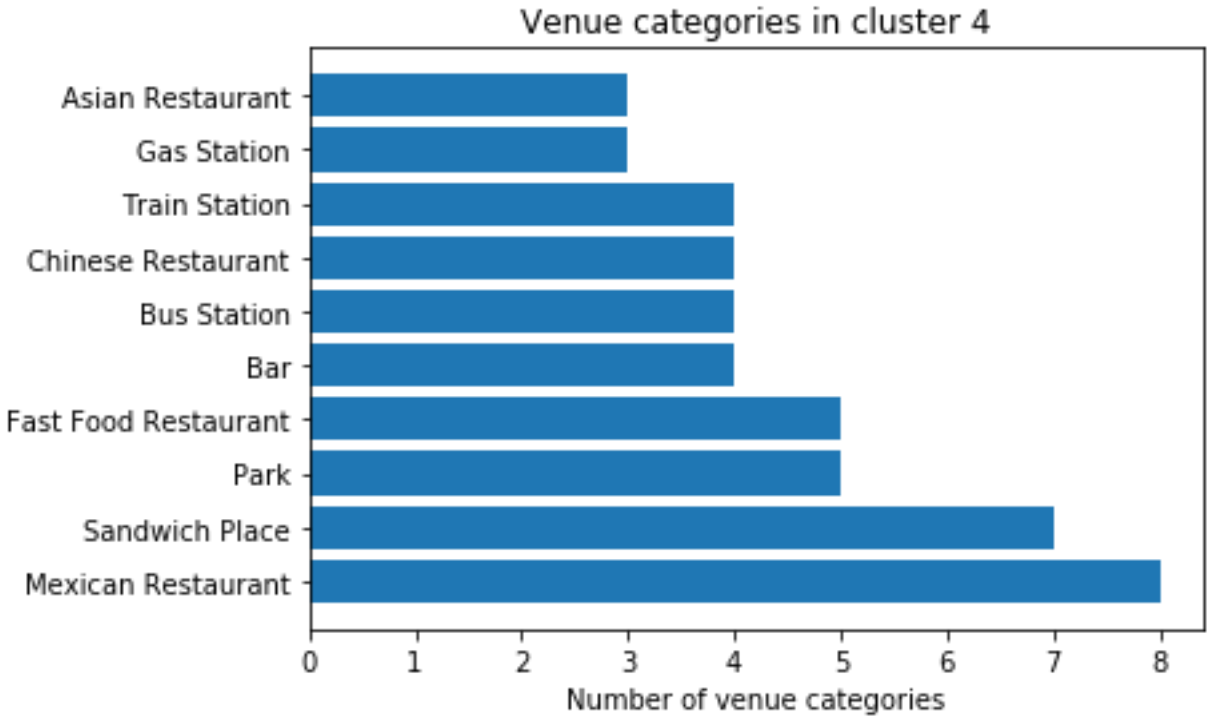
Cluster 1 consists of neighborhoods with bar , restaurants, ,supermarket, mobile phone shop and convenience stores .



Cluster 2 is a cluster of neighborhoods with restaurants, bakeries, clothing store, supermarket, mobile phone shop, coupled with pharmacy and train.



Cluster 3 has mostly sandwich places, train station, high end restaurants and hotels with bank and park, which suggests it is made up of downtown neighborhoods.



Venue categories in cluster 4 appear to be predominantly restaurants and fast food type restaurants which all suggest places where one can find something quick to eat, coupled with park, bus station and some gas stations.

Discussion

Among the above clusters it would appear that the following clusters are best suited for families with children:

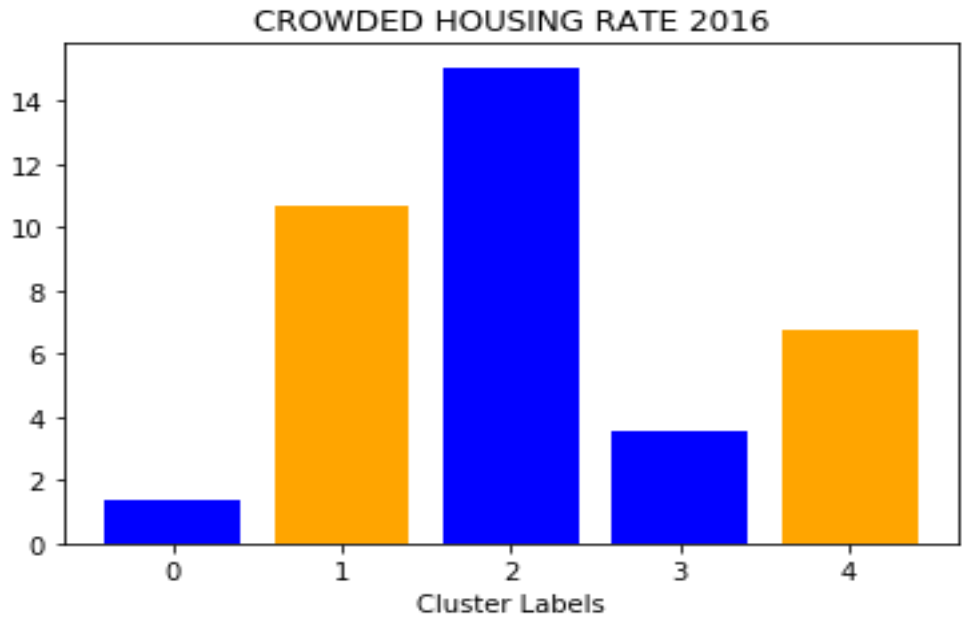
Cluster 2: Residential. Venue categories in this cluster are predominantly restaurants, bakeries, clothing store, supermarket, mobile phone shop, coupled with pharmacy and train.

Cluster 1: This is a cluster of neighborhoods with bar, restaurants, supermarket, mobile phone shop and convenience stores.

Some analysis was done to verify these observations.

The average percent of private dwellings occupied by usual residents in each cluster was checked to verify the results. The results are shown below.

The figure below shows that Clusters 2 and 1 that we have identified as best suited for families with children are obtained have a higher percent of housing crowded than other neighborhoods .This verifies our results.



Recommendation

The recommendation for families with children who are looking for a neighborhood to move to in Chicago would be to look for places in clusters 2 and 1.

Conclusion

As shown in the analysis here, the best place for families with children who are looking for neighborhoods in Chicago is to look for places in clusters 2 and 1.

The suggestions for future study can be the following

(1) Conduct additional analysis with private school data sets

(2) Look for additional data sets like employment , household income, public health facilities, hardship index etc for comparison

(3) Also find venue category/Crowd sourced data using websites like Trip Advisor along with FourSquare API

(4) Compare with other online sites for the best places to live in Chicago